THE EFFECTS OF INSTRUCTOR DIFFERENCES UPON STUDENT PROGRESS IN NAVAL AVIATION TRAINING

Wayne L. Waag, et al

Naval Aerospace Medical Research Laboratory
Pensacola, Florida

10 November 1972

NAMRL - 1171

THE EFFECTS OF INSTRUCTOR DIFFERENCES UPON STUDENT PROGRESS IN NAVAL AVIATION TRAINING

Wayne L. Waag, Ph.D.

and

LT Richard H. Shannon, MSC, USN

10 November 1972

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY

PENSACOLA FLORIDA

16 R

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Naval Aerospace Medical Research Laboratory
Naval Aerospace Medical Institute
Naval Aerospace and Regional Medical Center
Pensacola, Florida 32512

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
THE EFFECTS OF INSTRUCTOR DIFFERENCES UPON STUDENT PROGRESS IN NAVAL AVIATION TRAINING

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
Wayne L. WAAG, Ph.D.
LT Richard H. SHANNON, MSC, USN

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
10 November 1972	13	4

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO MF51.524.002.5012 DX5X	NAMRL 1171
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) .4
d.	

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The present investigation attempted to determine: (1) whether instructor differences could be measured quantitatively; (2) if such differences affected the grades which they assigned; and (3) if such differences affected the student's progress through the flight training program. Using an unstructured rating form, it was found that reliable instructor differences could be identified in terms of how they characteristically evaluate students. Furthermore, such differences were found to affect the grades which they assigned, although the magnitude of such effects was quite small. Moreover, these differences were not found to affect the student's progress through the program in terms of his pipeline assignment, subsequent flight grades, or his chances of receiving his wings. These data support the contention that flight instructor standardization procedures from an operational point of view have been successful.

DD FORM 1 NOV 68 1473 (PAGE 1)
S/N 0101-807-6801

Unclassified
Security Classification

I

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Pilot Performance Measurement						
Instructor Differences						
Aviation Training						
Standardization Procedures						

Approved for public release; distribution unlimited.

THE EFFECTS OF INSTRUCTOR DIFFERENCES UPON STUDENT PROGRESS IN NAVAL AVIATION TRAINING

Wayne L. Waag, Ph.D., and LT Richard H. Shannon, MSC, USN

Bureau of Medicine and Surgery
MF51.524.002.5012DXSX.4

Approved by

Ashton Graybiel, M.D.
Assistant for Scientific Programs

Released by

Captain N. W. Allebach, MC, USN
Officer in Charge

10 November 1972

Naval Aerospace Medical Research Laboratory
Naval Aerospace Medical Institute
Naval Aerospace and Regional Medical Center
Pensacola, Florida 32512

SUMMARY PAGE

PROBLEM

The purpose of this study was to answer the following questions. First, are there differences among primary flight instructors that can be measured quantitatively? If differences can be identified and measured, do they affect the actual grades which are assigned during Primary Flight Training? Do such differences affect the student's progress through the flight training program in terms of his subsequent flight grades, his chances of completing the program, or his pipeline assignment?

FINDINGS

It was demonstrated that instructor differences could be measured reliably by means of a relatively unstructured rating form. Differences in these ratings obtained across instructors were also reflected in the actual grades which were assigned. However, the magnitude of such differences was quite small. Furthermore, no differences were obtained across students assigned to different instructors in terms of their pipeline assignment percentages, their subsequent Basic and Advanced flight grades, or their attrition rate percentages. When certain instructors were selected and categorized as extremely "high" raters or "low" raters, similar results were obtained. In summary, it appears that instructor differences affect the grades which are assigned during primary flight training in a statistical sense but not in a practical sense since these differences do not affect the student's progress through the program or his subsequent flight performance grades. The data suggest that to a large extent flight instructor standardization procedures have been relatively successful.

INTRODUCTION

Most research concerning naval flight training personnel has focused upon the student. Many ability and performance configurations which discriminate between successful and unsuccessful student pilots have been identified. The present Student Pilot Prediction System attests to the success of these efforts (1). Within the flight program, however, the student aviator represents but one of the essential components. The training syllabus, the aircraft, and the instructor likewise affect the efficiency of the flight program. Data on these elements unfortunately are quite limited. The present investigation is concerned with the neglected personnel component--the instructor.

In most cases, Primary flight training represents the naval aviation student's first encounter with flying an aircraft. It is during this phase of training that his attitudes toward aviation are shaped and basic flying skills developed. Since the flight instructor must serve the dual role of teacher and evaluator, a concerted effort is directed toward his standardization. All incoming instructors are required to attend several weeks of indoctrination classes. They must also complete a flight phase consisting of 21 hops in which an attempt is made to develop a standardized method of flight instruction. Once the new instructor begins teaching, he must follow a standardized commentary for the introduction and demonstration of all flight maneuvers. The grades which he assigns are closely monitored in order to standardize their distributional characteristics. All of these measures are directed toward the reduction of any variability in the training system which could be attributed to instructor differences.

Despite these precautions, the potential for instructor differences still remains. First of all, instructors assigned to VT-1, the Primary flight training squadron, represents a relatively heterogenuous group. In a recent survey, approximately two-thirds were found to be "sergrads"; that is, recently designated pilots without any fleet experience (4). Of the remaining sample, approximately half were helicopter pilots while the other half was a mixture of pilots from the attack, fighter, patrol, and transport communities. Within the group of fleet-experienced pilots, there were vast differences in terms of the actual number of flight hours. Furthermore, at any given time, there are differences in terms of the length of duty as an instructor. For these reasons, it is apparent that some instructor differences will always exist. However, the nature and extent of these individual differences, and their effect, if any, upon student flight performance and evaluation are not known.

things--actual differences in flight ability or artifactual differences resulting from instructor variability.

It seems reasonable to assume that each instructor evaluates student pilot performance according to an internal frame of reference; that is, to some extent the standards he sets must influence his judgment of acceptable levels of performance. Due to the highly subjective nature of the instructor's "internal criterion", there is some evidence to suggest that it can best be measured by instruments which are highly unstructured (3). The grades an instructor assigns do not meet these requirements. He is told to maintain an overall average of 3.00 in which 20% of his ratings are "Below Average", 60%, "Average", and 20%, "Above Average".

In connection with the continuation of a student prediction study (2), an experimental rating form of student pilot performance had been completed by instructors for a large sample of student aviators in Primary training between July 1969 and December 1970. The raw data from this form were made available for this study. Specifically, instructors were asked after the 7th or 8th hop to rate their students on each of four questions concerning: (1) the probability of the student obtaining his wings; (2) the student's motivation; (3) the student's headwork; and (4) the student's reaction to stress. The complete questionnaire is presented in Appendix A. All questions were rated on a 13 point scale in which the anchor points were non-specific and highly subjective. Due to the lack of structure of the instrument, it was felt that the obtained responses would provide an adequate estimate of the instructor's "internal criterion". From the total 233 instructors, it was decided to include only those who had rated at least 15 students during this time period. A total of 70 instructors having 1330 flight students met this requirement. For each student, the ratings on the four items from the questionnaire and the flight grades from the PS, PCN, and TRANS stages of training were obtained. Pipeline assignment, the Basic flight grade, and the Advanced flight grade were also recorded. Furthermore, each student was categorized as a "completion" or "attrition". Of the 1330 students, 82 were in the later stages of advanced training and were consequently considered "completions" since attritions are negligible at these advanced phases. In order to determine whether the quality of students differed across instructors, certain selection test scores were also recorded. These included the Aviation Qualifying Test (AQT), the Spatial Apperception Test (SAT), the Mechanical Comprehension Test (MCT), the Biographical Inventory (BI), and the Flight Aptitude Rating (FAR), which is a weighted combination of the SAT, MCT, and BI.

RESULTS

To determine whether differences existed across instructors in terms of the quality of their students, each of the selection test scores was used as the dependent measure in a series of one-way analyses of variance. For each analysis, 70 treatment levels were defined--each comprised of the scores of all students assigned to an individual instructor. F-ratios of 1.135, 1.059, 0.948, 0.991, and 1.140 were obtained for the AQT, MCT, SAT, BI, and FAR, respectively. None of these values was statistically significant indicating no between-student differences across instructors. Consequently, if student performance differences emerged across instructors, it seemed highly unlikely they could be attributed to individual differences in the quality of students assigned to each instructor.

One-way analyses of variance were then performed for each of the four items from the questionnaire. F-ratios of 3.474, 5.501, 4.299, and 4.705 were obtained for the four items respectively. All values were highly significant ($p < .001$). In order to obtain estimates of the magnitude of instructor differences, a correlation ratio (corrected for shrinkage) was computed for each item. The obtained values were 0.337, 0.435, 0.382, and 0.401 respectively, indicating that from 11.36% to 18.92% of the variance of the ratings could be attributed to differences among instructors.

To ascertain whether such differences affected the actual grades the instructor assigned, similar analyses were performed using the PS and PCN grades as dependent measures. F-ratios of 1.673 and 1.580 were obtained for these two stages respectively. Both values were statistically significant ($p < .01$). Correlation ratios of 0.187 and 0 176 were obtained indicating that instructor differences accounted for only 3.48% and 3.09% of the variability of the grades respectively. It seemed apparent that the instructor differences reflected on the unstructured questionnaire were also evident in the grades which were assigned. The possibility remained, however, that such differences may have reflected differences in the quality of student flight performance. If such were the case, one would expect these differences to also be manifested in the next phase of training, the TRANS stage. Using the same analysis, an F-ratio of 1.189 was obtained--a value which is not statistically significant. These findings suggest that average differences in the ratings and grades during Primary flight training (PS and PCN) reflect differences in the instructor's "internal criterion" and not differences in actual student flight performance. The possibility remained that the absence of reliable differences during the TRANS stage may have occurred in the event that performance during Primary flight training was unrelated to flight performance during this later stage. However, correlations of .423 and .343 were obtained between the PS, PCN, and TRANS stage grades respectively, indicating them to be significantly related.

To determine whether instructor differences affected the student's later performance in training, analyses were performed using the Basic flight grade and the Advanced flight grade. F-ratios of 0.923 and 0.971 were obtained respectively indicating no differences. Perhaps the two most important events during undergraduate training for the naval aviator concern his pipeline assignment and whether or not he receives his wings. Since instructor differences were obtained for the grades which are assigned and since pipeline assignment is to a large extent based upon Primary flight training grades, it seemed likely that differences should also be reflected in terms of pipeline assignments. Students were categorized as entering either the jet or prop pipeline. All students assigned to the helicopter pipeline were included in the prop category since Basic training for those two pipelines is much the same. The relative frequency of students across instructors assigned to each pipeline was compared using X^2 No significant differences were obtained ($X^2 = 77.825$, df = 69). Finally, pass-attrite differences were tested across instructors. Likewise, no significant differences were obtained ($X^2 = 55.415$, df = 69). In summary, significant instructor effects were obtained for only the rating data and the grades assigned during Primary training. These findings are summarized in Table 2.

While no effects were obtained across the entire sample of 70 instructors, the possibility remained that differences might exist for "extreme" raters in the sample--that is, instructors who tended to give extremely high or extremely low ratings. To test for this possibility, two groups were defined--"high raters" and "low raters". Since ratings across the four items on the questionnaire were found to be highly intercorrelated (See Table 3), and since Item 1 (concerning the probability of the student securing his wings) was the most highly correlated with the pass/attrite dichotomy, the selection of instructors into the two extreme groups was based on the mean ratings for this item.

Population estimates of the mean and variance were computed for Item 1 using the entire sample. The standard error of the mean was computed using the mean number of students per instructor ($\overline{N} = 19$) as an estimate of sample size. An instructor was selected as a "high" or "low" rater if his mean rating was at least $\pm$ 2.58 standard errors above or below the population estimate of the mean. In other words, if z-tests had been performed for each instructor, comparing his mean rating with the population estimate based upon the entire sample, only those instructors were selected as "extreme" raters whose difference would have been significant beyond the .01 level. Using this rationale, 11 "high" rater instructors with a total of 220 students while 13 "low" rater instructors were selected with a total of 245 students. The mean rating of students assigned to "low" rater instructors was 7.227 as compared with 10.550 for students assigned to "high" rater instructors.

Table 2

Summary of Analyses of Variance for Performance Measures Across Instructors

Performance Measure	F Ratio
Aviation Qualifying Test	1.135
Mechanical Comprehension Test	1.059
Spatial Apperception Test	.948
Biographical Inventory	.991
Flight Aptitude Rating	1.140
Item 1--Wings	3.474**
Item 2--Motivation	5.501**
Item 3--Headwork	4.299**
Item 4--Stress	4.705**
Pre Solo Grade	1.673*
Precision Grade	1.580*
Transition Grade	1.189
Basic Flight Grade	.923
Advanced Flight Grade	.971
Pipeline Assignment	77.825+
Pass/Attrite	54.415+

**p < .001
*p < .01
+X^2 Value

Table 3

Intercorrelations Among Ratings and Pass/Attrite

	1	2	3	4	5
1. Item 1-Wings	1.000	.692	.799	.786	.278
2. Item 2-Motivation		1.000	.648	.622	.204
3. Item 3-Headwork			1.000	.867	.231
4. Item 4-Stress				1.000	.245
5. Pass/Attrite					1.000

For each of the dependent measures, z-tests were performed comparing these two groups. The results are presented in Table 4. As indicated, only the PS grade was found to be significant. The grades of the students assigned to "upper" rater instructors were significantly higher than grades of students assigned to "lower" rater instructors. A point-biserial correlation coefficient was computed and found to be 0.094 indicating the "high" rater—"low" rater dichotomy to account for only 0.89% of the variance of the PS grades.

DISCUSSION

The results of this investigation clearly indicate that differences across primary flight instructors can be measured quantitatively. For the ratings obtained in response to the relatively unstructured questionnaire, instructor differences accounted for approximately 11% to 19% of their variability. The results furthermore suggest that such differences affect the grades which the instructor assigns. However, the magnitude of such effects is substantially reduced. In fact, instructor differences accounted for only 3% of the variance of the PS and PCN grades. When certain instructors were classified as extreme "high" or "low" raters, the amount of explained variance was reduced

Table 4

Comparison of Performance Measures Between Students Assigned To "High" Rater and "Low" Rater Instructors

Performance Measure	Means "High" Rater	"Low" Rater	z Value
Aviation Qualifying Test	85.870	86.148	.258
Mechanical Comprehension Test	59.824	59.524	.424
Spatial Apperception Test	21.524	21.775	.469
Biographical Inventory	39.588	40.515	.815
Flight Aptitude Rating	6.185	6.189	.032
PS Grade	3.030	3.018	1.960*
PCN Grade	3.066	3.061	.652
TRANS Grade	3.008	3.005	.515
Basic Flight Grade	3.030	3.026	.861
Advanced Flight Grade	3.052	3.053	.078
Pipeline Assignment	.370	.376	.128
Pass/Attrite	.754	.722	.788

*$p < .05$

to less than 1%. Such data suggest that present standardization efforts are to a large extent successful in reducing inter-instructor variability in student flight performance evaluations.

Likewise, instructor differences were found to have little effect upon the student's progress through the program. Contrary to popular belief--especially among flight students--assignment to a "high" rater or "low" rater instructor had no effect upon their subsequent pipeline assignment. Furthermore, no differences were reported across instructors in terms of student flight performance as measured by the Basic and Advanced flight grades. Of greatest importance, no statistically reliable differences were reported for the pass/attrite percentages.

The results of this study suggest that while instructor differences can be isolated and quantitatively measured, their effect is quite negligible in a practical sense. The reduction of the variability attributable to instructor differences for the actual grades which are assigned attests to the success at attempts toward standardization. Such findings are consistent with previous

evidence suggesting that raters can be trained to reduce internal sources of bias (3). That instructor differences had no effect upon the student's subsequent flight performance, his pipeline assignment, or his chances of receiving his wings is highly encouraging. It suggests that the present concern with the effects of instructor differences may be unwarranted.

It should be strongly emphasized that the findings of this investigation are based upon a relatively large sample of instructors. Although the data indicate that, on the basis of those sampled, instructor differences are relatively unimportant, this finding does not guarantee that certain individuals could well deviate substantially from those in this study. In an ever changing instructor population, it therefore seems prudent to concede the possible presence of a few individuals who could adversely affect their student's progress through the flight training program. It remains the responsibility of the training command to monitor the performance of its instructors in order to identify such deviant individuals and subsequently modify their teaching behavior.

REFERENCES

1. Aerospace Psychology Department, Manual for Use of the Student Pilot Prediction System. Naval Aerospace Medical Research Laboratory, Pensacola, Florida.

2. Berkshire, J. R., and Ambler, R. K. The value of indoctrination flights in the screening and training of naval aviators. Aerospace Medicine, 1963, 34, 420-423.

3. Bergman, B. A., and Siegel, A. I. Training evaluation and student achievement measurement: A review of the literature. AFHRL-TR-72-3. Lowry AFB, Colorado: Air Force Human Resources Laboratory, 1972.

4. Long, G. M., Shannon, R. H., and Waag, W. L. The isolation of critical elements within selected maneuvers during primary flight training. Proc. 16th Annual Meeting of Human Factors Society, 1972, 385-388.

APPENDIX A

INSTRUCTOR'S RATING

Instructor's name _____

Student's name _____

Jacket number _____

What is the last hop this student completed? _____

Studies have shown that primary flight instructors are in the best possible position to make an early evaluation of an individual student. Such an early assessment would be a valuable addition to the information administrators now have available when evaluating a student. This questionnaire will not be kept in the student's jacket but in a separate file.

Below are four questions for you to answer. The questions are subjective and are difficult to answer definitively. To get an accurate assessment of your opinion please check the line on the continuum which best represents your feeling.

1) IN YOUR OPINION WILL THIS STUDENT GET HIS WINGS?

definite no	probably will	definite yes

2) HOW WELL MOTIVATED IS THIS MAN TO BECOME A NAVAL AVIATOR?

not very	well motivated	extremely well

3) HOW IS THIS STUDENT'S HEADWORK?

poor	good	outstanding headwork

4) HOW MUCH CONTROL DOES THIS MAN HAVE WHEN UNDER STRESS?

poor	good	outstanding control